# AND/OR Branch-and-Bound for Computational Protein Design Optimizing K*

Bobak Pezeshki (pezeshkb@uci.edu), Radu Marinescu (radu.marinescu@ie.ibm.com), Alex Ihler (ihler@ics.uci.edu), Rina Dechter (dechter@ics.uci.edu)

UC Irvine

## Summary

We introduce AOBB-K*MAP, a new branch-and-bound algorithm over AND/OR search spaces for solving the K*MAP protein design problem. In addition to formulating CPD as a graphical model for K* optimization and providing a new efficient algorithm, we also introduce a statically compiled heuristic for K*MAP not previously used in CPD. This work extends algorithms for Marginal MAP (MMAP) and provides a framework and proof-of-concept for continued adaptation of existing state-of-the-art mixed inference schemes over AND/OR search spaces to protein design.

## Formulation

$\mathcal{M} = < X, D, F >$

$X = R \cup C_\gamma$

Two sets of variables:

- *R*esidue variables – capture the amino acid (aa) at the respective position in the protein
- *C*onformation variables - side chain conformation of the aa at the corresponding residue for protein states $\gamma \in \{\text{Bound}, \text{Unbound}\}$

$D = D_R \cup D_{C_\gamma}$

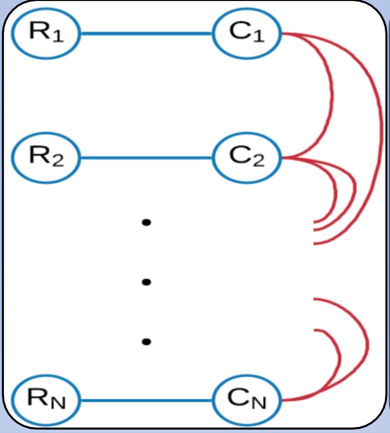
Domains:

- $D_R$ - amino acids being considered at each position
- $D_{C\gamma}$ - various side chain conformations for all aa considered at the corresponding residue for each protein state γ

$F = E_\gamma \cup \mathcal{C}$

Functions:

- *E*nergy of interactions between the amino acids for each protein state γ
- *𝒞*onstraints enforcing consistency between side chain conformations and amino acid assignments



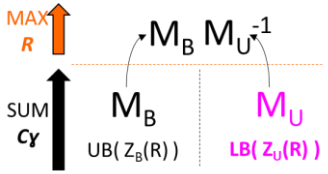

## K* Objective

$$K^*(R_1, \dots, R_N) = Z_B(R_1, \dots, R_N) / Z_U(R_1, \dots, R_N)$$

$$Z_\gamma(R_1 \dots R_N) = \sum_{C_1,\dots,C_N} \prod_{\mathscr{C}_{\gamma(i)} \in \mathscr{C}} \mathscr{C}_{\gamma(i)}(R_i, C_{\gamma(i)}) \cdot \prod_{E^{sb}_{\gamma(i)} \in \boldsymbol{E}^{sb}_\gamma} e^{-\frac{E^{sb}_{\gamma(i)}(C_{\gamma(i)})}{\mathscr{R}T}} \cdot \prod_{E^{pw}_{\gamma(ij)} \in \boldsymbol{E}^{pw}_\gamma} e^{-\frac{E^{pw}_{\gamma(ij)}(C_{\gamma(i)}, C_{\gamma(j)})}{\mathscr{R}T}}$$

$$task: \; K^*MAP = \max_{R_1,\dots,R_N} K^*(R_1, \dots, R_N)$$

## MBE-K*




Based on a constrained ordering such that conformation variables are processed first, messages corresponding to upper and lower bounds on the partition functions of the conformation variables associated with the various protein states are computed and passed up to be maximized over

**Algorithm 2:** AOBB-K*MAP

**input** : CPD graphical model $\mathcal{M}$; pseudo-tree $\mathcal{T}$; $K^*$ upper-bounding heuristic function $h^{ub}_{K^*}(.)$; $Z_\gamma$ upper-bounding heuristic function $h^{ub}_{Z_\gamma}(.)$; and subunit stability threshold $threshold(\gamma)$ for each subunit $\gamma \in \varphi$

**output**: $K^*MAP(\mathcal{M})$

**begin**
  Initialize $MiniSat$ with constraints from $\mathcal{M}$ and generate literals via constraint propagation
  $\pi \leftarrow$ dummy AND node $n_D$
  $ub_{K^*}(n_D) \leftarrow \prod_{m \in ch_\mathcal{T}(n_D)} h_{K^*}(m)$
  $lb_{K^*}(n_D) \leftarrow -inf$
  $g(n_D) \leftarrow 1$
  **foreach** $\gamma \in \varphi$ **do**
    $A^\times_{Z_\gamma}(n_D) \leftarrow 1$
    $A^+_{Z_\gamma}(n_D) \leftarrow 0$
    $ub_{Z_\gamma}(n_D) \leftarrow \prod_{m \in ch_{\mathcal{T}_\gamma}(n_D)} h_{Z_\gamma}(m)$
  **while** $n_X \leftarrow EXPAND(\pi)$ **do**
    **if** $MiniSat(\pi) = false$ **then**
      $PRUNE(\pi)$
    **else if** $\exists \gamma \in \varphi$ *s.t.* $UB_{Z_\gamma}(n_X) < threshold(\gamma)$ **then**
      $PRUNE(\pi)$
    **else if** $X \in \boldsymbol{R}$ **then**
      **if** $\exists a \in anc^{OR}(n)$ *s.t.* $ub_{K^*}(a, \pi) < lb_{K^*}(a)$ **then**
        $PRUNE(\pi)$
      **end**
    **else if** $ch^{unexp}_\mathcal{T}(n) = \emptyset$ **then**
      $BACKTRACK(\pi)$
  **end**
  **return** $ub_{K^*}(n_D) = lb_{K^*}(n_D) = K^*MAP(\mathcal{M})$
**end**

| Problem | iB | \|X\| | Dmax | w* | d | UB | OR | AND | CPP | UBP | SSP | EH | time | K*MAP | BBK* t | BBK* sln |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1gwc_00021 | 4 | 12 | 203 | 4 | 6 | 10.29 | 28766 | 134930 | 77823 | 55 | 2 | 5 | **16** | 9.79 | 152 | 9.79 |
| 2hnu_00026 | 4 | 14 | 203 | 5 | 7 | 15.08 | 22010 | 105458 | 76657 | 38 | 0 | 4 | **7** | 13.18 | 437 | 13.18 |
| 2hnv_00025 | 4 | 16 | 203 | 6 | 8 | 15.04 | 115194 | 297138 | 84882 | 39 | 0 | 3 | **16** | 13.65 | 962 | 13.65 |
| 2rf9_00018 | 6 | 18 | 205 | 7 | 9 | 16.68 | 20137 | 85033 | 87306 | 78 | 0 | 4 | **15** | 15.79 | 187 | 15.79 |
| 2rfd_00035 | 6 | 16 | 205 | 6 | 8 | 17.70 | 896239 | 4253159 | 3273123 | 40 | 0 | 4 | **381** | **17** | 1242 | 16.77 |
| 2rfe_00030 | 4 | 14 | 203 | 5 | 7 | 11.53 | 20393 | 164126 | 359007 | 87 | 40 | 7 | **19** | 10.50 | 182 | 10.50 |
| 2rfe_00043 | 6 | 16 | 203 | 6 | 8 | 18.48 | 15390 | 40297 | 422357 | 34 | 43 | 4 | 80 | 18.04 | 50 | 18.04 |
| 2rfe_00044 | 6 | 16 | 203 | 6 | 8 | 18.62 | 37887 | 99927 | 1047107 | 30 | 3 | 6 | 86 | 18.19 | 75 | 18.19 |
| 2rl0_00008 | 4 | 10 | 203 | 3 | 5 | **11.16** | 2 | 3 | 0 | 40 | 0 | 3 | **3** | **11.16** | 262 | 9.46 |
| 2xgy_00020 | 4 | 14 | 203 | 5 | 7 | 11.47 | 43643 | 262523 | 743860 | 40 | 0 | 2 | **14** | 10.60 | 887 | 10.60 |
| 3cal_00032 | 6 | 16 | 203 | 6 | 8 | 13.38 | 133851 | 1067419 | 531976 | 32 | 6 | 4 | **125** | 11.62 | 1429 | 11.62 |
| 3u7y_00009 | 5 | 12 | 203 | 4 | 6 | **4.51** | 2 | 3 | 0 | 40 | 0 | 3 | **6** | 4.51 | 191 | 4.51 |
| 4kt6_00023 | 4 | 16 | 203 | 6 | 8 | 14.80 | 38186 | 101546 | 23877 | 16 | 19 | 4 | **7** | 12.69 | 136 | 12.69 |
| 4wwi_00019 | 5 | 14 | 203 | 5 | 7 | 15.43 | 8094 | 30774 | 17888 | 40 | 0 | 2 | **7** | 14.99 | 26 | 14.99 |
| 1gwc_00021 | 4 | 13 | 203 | 4 | 7 | 12.51 | 33881 | 590621 | 473189 | 388 | 6 | 8 | **205** | **11.92** | 551 | 11.72 |
| 2hnv_00025 | 4 | 17 | 203 | 6 | 9 | 18.38 | 215171 | 550559 | 220825 | 77 | 0 | 4 | **153** | **16.18** | 880 | 13.65 |
| 2rfe_00012 | 5 | 15 | 205 | 5 | 8 | 14.36 | 3127 | 10003 | 32610 | 57 | 0 | 3 | 85 | 13.93 | 12 | 13.93 |
| 2rfe_00014 | 5 | 15 | 205 | 5 | 8 | 14.79 | 4087 | 13087 | 39411 | 57 | 0 | 3 | 85 | 14.36 | 45 | 14.36 |
| 2rfe_00017 | 5 | 15 | 203 | 5 | 8 | 11.46 | 245894 | 1063198 | 6389737 | 227 | 25 | 43 | 333 | **10.86** | 78 | 10.80 |
| 2rfe_00030 | 4 | 15 | 203 | 5 | 8 | 13.61 | 256957 | 1327425 | 2816050 | 726 | 83 | 77 | **274** | **11.12** | 275 | 10.97 |
| 2xgy_00020 | 5 | 15 | 203 | 5 | 8 | 11.39 | 398102 | 2383318 | 7422285 | 42 | 0 | 20 | **360** | 10.96 | 1388 | 10.96 |
| 3u7y_00009 | 4 | 13 | 203 | 4 | 7 | 4.95 | 36760 | 228568 | 564654 | 204 | 7 | 28 | **99** | 4.51 | 216 | 4.51 |
| 3u7y_00011 | 4 | 13 | 203 | 4 | 7 | 12.29 | 5758 | 16108 | 68579 | 50 | 0 | 5 | 86 | 11.85 | 27 | 11.85 |
| 4wwi_00019 | 5 | 15 | 203 | 5 | 8 | 16.05 | 22945 | 87485 | 91677 | 176 | 75 | 5 | 180 | 14.99 | 34 | 14.99 |

## Results and Future Work

**Preliminary Results**:

- AOBB-K*MAP performs better than BBK* time-wise for the majority of problems tested with 2 MAP variables and better than half of those with 3 MAP variables
- wMBE-K* provides solutions to conditioned subproblems during search, sometimes able to solve the entire K*MAP problem exactly
- AOBB-K*MAP finds solutions with greater K* values than BBK*

**Future Work**:

- Incorporation of constraints into wMBE-K* to exploit determinism, and the use of a dynamic heuristic
- Adaptation of state-of-the-art [approximate] search and sampling algorithms to solving the K*MAP problem.